DOCUMENT RESUME

ED 064 401 TM 001 649

AUTHOR Gable, Robert K.; Pruzek, Robert M.
TITLE Methodology for Instrument Validation: An Application to Attitude Measurement.
PUB DATE Apr 72
NOTE 26p.; Paper presented at the annual meeting of the AERA (Chicago, Ill., April 1972)

EDRS PRICE MF-$0.65 HC-$3.29
DESCRIPTORS Evaluation Methods; Factor Analysis; Item Analysis; *Measurement Instruments; Negroes; *Racial Attitudes; *Rating Scales; Test Construction; *Test Validity

ABSTRACT

The purposes of this paper are to (1) discuss general issues related to a newly proposed methodology for item content and construct validation and (2) apply the new method in the context of developing a scale for measuring attitudes toward black people. Latent partition analysis was proposed and applied for the study of item content validity using partitioned data from 36 college presidents. Factor analysis was used to assess construct validity using item response data from 212 college students. A comparison between the item response factors and the LPA categories indicated substantial agreement. Limitations and virtues of the present methodology are discussed in relation to certain other methods for content and construct validation. (Author)

ED 064401

12.9

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

Methodology for Instrument Validation: An Application to Attitude Measurement

Robert K. Gable
University of Connecticut

Robert M. Pruzek
State University of New York at Albany

TM 001 649

FILMED FROM BEST AVAILABLE COPY

Paper presented at the American Educational Research Association annual meeting, April, 1972, Chicago, Illinois.

ABSTRACT

The purposes of this paper are to (1) discuss general issues related to a newly proposed methodology for item content and construct validation and (2) apply the new method in the context of developing a scale for measuring attitudes toward black people. Latent partition analysis was proposed and applied for the study of item content validity using partitioned data from 36 college presidents. Factor analysis was used to assess construct validity using item response data from 212 college students. A comparison between the item response factors and the LPA categories indicated substantial agreement. Limitations and virtues of the present methodology are discussed in relation to certain other methods for content and construct validation.

Methodology for Instrument Validation: An Application to Attitude Measurement

Robert K. Gable
University of Connecticut

Robert M. Pruzek
State University of New York at Albany

Introduction

In the context of either attitude or achievement test development, two fundamentally different kinds of data are needed if derived scores are to be given valid interpretations: (1) judgments of subject-matter, or content, specialists regarding characteristics which items or tests are seen to measure, and (2) responses of individuals for whom the items or tests are considered to be appropriate.

One purpose of this paper is to examine both judgmental and response data which were gathered in the process of developing an instrument for measuring attitudes toward black people (Simons, 1971). Latent partition analysis (Wiley, 1967, Psychometrika) was used to study judgmental data gathered from 36 presidents of colleges which are members of the United Negro College Fund. Factor analysis was used to examine item response data from 212 university students who were enrolled in black studies courses.

A more egregious purpose is to discuss general strategy for developing tests. While the data for the present study are associated with attitude measurement, the discussion is relevant for the measurement of achievement as well as attitudes.

Classificatory Methods for Studying Content of Tests

Unlike the traditional methods for studying response data, methods for describing rationally the content of attitude (or achievement) tests have not been thoroughly studied. The notion of examining item content using classificatory methods has

been implicit in writings of several measurement specialists (e.g. Lindquist et al, 1950), but not many empirical studies of such methods have been published. In the few reported studies of content validity, evidence for content interpretations has typically been in the form of descriptive statistics indicating how adequately items fit a priori content categories.

Robert Ebel published a paper in 1953 in which a method was described for classifying items into named categories where the categories distinguished different types of cognitive abilities. For this work, trained measurement personnel of the Examinations Services at the State University of Iowa regularly classified items from classroom tests into what Ebel termed "relevance categories." The practice was carried on simply for the purpose of describing test contents and no detailed analysis of these classifications has been reported by Ebel or his colleagues.

A more systematic study of this type was carried out by McGuire (1963) in a study of certifying examinations in medical education. McGuire had several subject-matter specialists categorize items from medical certifying examinations into categories roughly equivalent to the levels of Bloom's (1956) taxonomy of the cognitive domain. Results were presented simply in terms of descriptive statistics of agreement among the various categorizations. Certain agreements were found among the categorizations but some substantial discrepancies were also noted. McGuire also employed a factor analysis to analyze correlations between the eight subtests defined (approximately) by her categorizations of items, but most correlations were low and derived factors were not interpreted.

A more sophisticated classificatory study of achievement test items was that of Stoker and Kropp (1964). These authors also used a categorical system derived from the Bloom taxonomy. Items were devised to correspond to the major levels of the taxonomy and then persons familiar with the various levels were

asked to sort items independently into the respective named categories. Explicit attention was given to the complexity continuum on which Bloom's categories are ordered. As in McGuire's study, results of the independent categorizations were reported simply in terms of descriptive statistics of agreement, and again, numerous discrepancies were found between resultant categorizations. Subsequently, relationships between results of the categorizations and results of response data analysis were discussed briefly.

Although the classificatory procedures discussed above surely constitute what may be termed ad hoc methods, the problems raised by the studies are significant. We shall consider such problems in the following section.

An Analytic Method for Item Content Validation

Measurement specialists typically do not regard content validity as primarily a quantitative characteristic of tests. Rather, this type of validity seems more properly viewed as qualitative. It is for this reason that methodology for item content validation may properly involve item classifications or categorizations. In this part of the paper we shall be concerned with the description of a method called latent partition analysis (LPA) for specifying content validity of attitudinal measures. Although we focus on LPA in this paper, it should be emphasized that our proposal rests more on the logic of a strategy of gathering data, and making decisions directly with respect to this data, then with any particular feature of LPA itself. The objectives and limitations of the proposed approach will be briefly considered in relation to features of typical methods of instrument development.

Latent partition analysis was formulated in connection with research[1] in which a method was needed to study relationships between different sorters' categorizations of the same set of items. For that research the items were statements

[1]Project No. 5.1015.2.12.1 (1967), supported by the U. S. Office of Education; Principal Investigator, Donald M. Miller, University of Wisconsin.

pertaining to teacher facilitation of learning in the elementary classroom which had been obtained through interviews with elementary teachers. Following the interviews, each of a number of teachers were given the same set of statements and were asked to sort these into what they considered homogeneous categories. Latent partition analysis (Wiley, 1967) was formulated and applied to study relationships among the resulting sets of categories (partitions).

Stated in Guttman's (1950) terms, the critical problems of building internally consistent tests are to define properly the intital universe of content and to select properly items for this universe. Guttman suggests that this should be a logical (deductive) process, but there are some formidable problems with the approach he describes. It is generally very difficult to know just what constitutes a particular universe, when two or more universes may be merged and when a universe should be broken down into subuniverses. In addition, once a universe has been defined, it may be difficult to generate items which in fact measure that which has been specified.

In order to circumvent the above problems, or at least to handle them more objectively, it is proposed that an inductive approach be used in generating internally consistent tests (or subtests). The specific suggestion is to begin from a pool of attitude (or achievement) items, to have each of several qualified persons sort the items into what they deem as homogeneous categories and to use an analytical method such as latent partition analysis (LPA) or hierarchical cluster analysis, to derive a single set of homogeneous item clusters. We propose that under certain conditions at least, content labels might be inductively derived from such a process; the items or labels can be modified at subsequent stages of research in order to refine an instrument.

Suppose that several subject-matter specialists were asked to sort independently a set of attitude (or achievement) items into homogeneous categories. Let us assume that no prior (named) categories have been specified. One would expect different persons to construct different categories; nevertheless, if each sorter had been given the same instructions for creating his categories, one would expect to find relationships among resulting categorizations. It is proposed that LPA be used to examine such a set of categorizations with the objective of discovering underlying agreements about contents which might not otherwise be evident. If each sorter were asked to generate categories on the basis of the content characteristics of the items, it seems reasonable to expect derived categories to reflect content characteristics of the items. That is, items within derived categories should be relatively more homogeneous in content than items between categories. Tucker (1962) noted that the critical problem in establishing content validity for achievement (or attitudinal) test items is specification of the universe of content for which items may be regarded as valid measures. Different qualified persons may be found not to agree about the particular universe (or universes) for an item. As described above, LPA or a similar method such as Johnson's (1967) hierarchical cluster analysis, may be used operationally to define content universes for items

In his chapter entitled Test Validation, Cronbach (1970, P. 446) essentially notes that the principal question of content validity is: "Do the observations [items] truly sample the universe of tasks [attitudes] the developer intended to measure?" The LPA method approaches this question by asking: "What are the constituent subuniverses for a universe which may be inductively derived from constructed items?" By examining derived LPA categories, and their mutual interrelationships, subuniverses are identified; then the subuniverses are studied in relation to the investigator's initial conception of the universe; items are added to or taken away from the original item pool on the basis of the above analysis.

The most important feature of the proposed approach is its relatively heavy reliance on objective methods for the finding and naming of content categories. Unlike most methods for classification, such as those reviewed, the present strategy involves an objective solution for item categories, even when the investigator has no named categories at the onset. While we agree with Rozeboom (1966) that our whole notion of content validity may require revision, and that the term "content" simply isn't a theoretical term, our approach seems at least to provide a better framework than most methods for ultimately interpreting item response data; this is our primary goal.

Methods for Studying Constructs of Tests

While assessments of content validity are based more or less exclusively on judgmental data, construct validity is based on examinations of judgmental data and of response data. Cronbach (1970) provides an enlightening discussion of construct validation as a continuing process of refining and elaborating on test (score) interpretations. In general, the process can be viewed in the context of three procedures: labeling hypothetical constructs, formulating testable hypotheses on the basis of some underlying theoretical network, and gathering evidence to test the hypotheses.

The initial labeling of the constructs consists of employing a method such as LPA to generate detailed descriptions of item content characteristics in the manner previously described. Following this, both confirmatory and exploratory approaches to factor analysis can be employed to answer the question: "Are factors which have been hypothesized in advance using judgmental data sufficient to account for the stable interitem covariation of the item response data?" Note the emphasis on relationships between content and construct validition. As a rule, when item responses have been influenced substantially by characteristics of the items which were not weighted heavily

in the judgmental data study one may expect discrepancies between LPA categories and response data factors. It is the detailed analysis of the nature and extent of such discrepancies which seems ultimately to be likely to add to knowledge of the constructs under study. While item response data may in general correspond to mere conceptualizing habits of a culture or to accidents of simultaneous learning of different constructs, and judgmental data categories may in turn be influenced by capriciousness of judges' perceptions, we believe that there is capital in making discrepancies explicit between these two fundamentally different kinds of information. Careful sampling of both judges and respondants will of course be most important in making studies such as these productive.

At this point we illustrate the suggested approach to instrument development in the context of a study by Simons (1971) which deals with the development of an instrument to measure attitudes toward black people.

LPA for Content Specification: An Empirical Study

In this section an empirical study by Simons (1971) which employed LPA to examine the content validity of an attitude instrument will be described.

Procedures

The categorical data examined in this section resulted from having 36 content specialists independently sort 51 items into categories. The items sorted reflected an item pool which had been generated on the basis of information obtained from literature reviews including journals, newspapers, magazines, radio, television, and lectures. Preliminary item screenings and data analyses by graduate students suggested that 51 items were worthy of further study.

The 36 sorters were presidents of colleges and universities located in Southern and border state that were members of the United Negro College Fund. The sorters were selected because of their association with faculty, alumni, and student groups which were predominantly black.

Items were put on individual slips of paper (without item numbers) and mailed along with instructions to sort the items into from 6 to 15 mutually exclusive and jointly exhaustive categories on the basis of similar item content.

Judges' categories, called manifest categories, were used as the starting point for latent partition analysis (LPA). The LPA program computes a joint proportion matrix (of order 51 x 51) where each entry indexes the proportion of (36) sorters who placed a given pair of items in the same manifest category. From this matrix, a latent category matrix is derived; entries in this matrix index for each item the derived (latent) category to which the item belongs.

Results of the LPA Study

Table 1 presents the latent category matrix; rows represent items and columns, categories. Inspection of the entries in Table 1 indicates that 12 latent categories were derived.

Insert Table 1 About Here

and that, for the most part, items are of unit complexity. Thus, we regard the LPA model as reasonable for these data. The following is a category-by-category description of each category in terms of what we call item content. Although we have taken the liberty of briefly naming each category, the reader should recognize that some liberties have been taken and that subsequent tables (especially 2 and Appendix A) ought also to be examined.

Category 1 was called Racism, since the item content states that black people have low moral standards and are innately inferior. Category 2 was named Black Studies Courses; all items in this cluster described various aspects of the objectives and characteristics of such courses. The third category, Inner City Teachers, was very clearly defined by items describing characteristics of inner city teachers.

Category 4 was called Public Opinion as all three items referred to the views of the university community toward blacks. The next category, 5, was labeled Community Control; it was neatly defined by items suggesting community control of education, police and business. Also clearly defined was Category 6, Philosophy. Items in this category dealt with the credibility of such organizations as the NAACP, SCLC and CORE.

Items reflecting characteristics of black students defined Category 7, Student Image. Category 8, was termed Cultural Differences since the item content suggested the existence of black art, theater and music. Although Category 9 was called Race Interaction, the three items defining the category were deleted from the study. Several judges commented that the items were inaccurate and potentially offensive.

Category 10, Personal Judgment, was defined by three items, one of which was offensive and thus deleted. The remaining two items reflected awareness or judgments of other people due to their skin color. Category 11 was called Cultural Background but was not clearly defined. Two of the three items defining the category were deemed offensive; the third item reflected possible teacher prejudice where a black student's classroom behavior tends to be rated low even though his academic achievement is satisfactory. Finally, Category 12 was called Bigotry since the items reflected the view that black students are inarticulate and apathetic.

The latent category matrix itself suggested that several clearly defined subuniverses can be identified for the judgmental data. Table 2 contains the matrix of indices of association between the pairs of latent categories. Inspection

Insert Table 2 About Here

of the entries suggested that some content categories reflected similar subuniverses of item content. Given this situation it was reasonable to combine some of the item categories to

produce a more parsimonious system of content clusters. Thus the following three sets of categories were examined for possible combination: Racism (1, Racism and 12, Bigotry), Public Opinion (4, Public Opinion and 10, Personal Judgment), and Black Studies (2, Black Studies and 8, Cultural Differences). Examination of the three categories indicated that the first two were quite consistent, but the tendency of the judges to combine the Cultural Differences items with Black Studies merely suggested the judges association of black art, theater and music with possible course topics in Black Studies. Since the instrument being developed was designed to reflect attitudes towards black art, theater and music, the Black Studies and Cultural Difference categories were not combined.

Thus our LPA results were taken to reflect eight categories of item content for the 51 items. Forty-one of the 51 items contributed to the naming of the categories in the latent category matrix. A total of six of these items were deemed inaccurate and possibly offensive by the judges and the researcher. To increase the internal consistency reliabilities of the scales on the instrument under development, 15 new items were written for selected categories to generate a 50 item instrument. Table 3 contains a summary of the original and final category labels and illustrates where

Insert Table 3 About Here

items were deleted or added on the basis of the LPA study. The original derived categories are designated LC's 1-12; the final judgmental categories are referred to as FJC I to FJC VIII. This notation will be used to distinguish between original and final categories.

In summary to this point, the LPA procedure provided a rational-inductive identification of the constituent sub-universes of item content. Examination of the derived categories, the relationships between categories, and the intended conception of the universe of content suggested certain item

deletions and additions with respect to the original item pool. A comparison between the eight LC's which were empirically generated on the basis of judgmental data, and actual item response factors will add further information regarding interpretation of the attitude constructs. Such a comparison will be presented in the next section.

A Factor Analysis of Response Data

After analysis of judgmental data, and suitable revisions of attitude items, an instrument comprised of 50 attitude items using a five point Likert format was administered to 212 university students who were enrolled in a black studies programs, or other courses designed to relate to this program. Some of the 50 items had been written to have negative stems while others were positive; all negatively stated items were (arbitrarily) reflected in the scoring.

Using these data a 50 x 50 matrix of intercorrelations was computed from which we generated principal components and followed with a normal varimax transformation. These results are presented in the following section with a description of relationships between the response data factors and the LC's from the LPA study. While we do not in this context present detailed analyses of the attitude constructs, an analysis is available in Simons (1971) and Simons and Gable (1972). (We also analyzed the 50 x 50 matrix using a modern version of scale-free image analysis but the factor interpretations were virtually identical to those of the derived components solution so we have chosen to include only the components results here.)

Results of Factor Analysis Study

Sixteen factors (components), accounting for 66% of the total variance were derived; nine of these were defined by at least two items with substantial loadings and, we therefore present a nine factor solution. The factor loading

matrix can be found in Appendix A. Table 4 contains the factor names, LPA original category (LC's 1-12) codes, LPA item numbers (for comparison with Table 1), the 50-item version item numbers, item summaries and factor loadings. Inspection of the entries in Table 4 indicates the extent that the item response factors reflect LPA item content categories.

Insert Table 4 About Here

Factor I was called Racism since it was defined by six of the eight LPA racism items from LC. Note that these six Racism items consisted of two items from the original Racism category, two from the original Bigotry category and two new Racism items. People tending to agree with the content of the items defining this factor might be considered to have racist attitudes.

Factor II was labeled Inner City Teachers; it was clearly defined by the seven Inner City Teacher items from FJC III. Recall that Table 3 indicated that FJC III consisted of five original Inner City Teacher items, one new item and one item from the original FJC II, Cultural Background. All seven of these items, grouped on the basis of judgmental data in the LPA study, were sufficiently interrelated for the student response data to generate Factor II. Agreement with the content of these items by a respondant suggests understanding and support of inner city teachers.

Factor III, Cultural Differences, was defined by all five of the items judged to form FJC VIII, Cultural Differences. Agreement with the item content acknowledges awareness of such cultural aspects of the black experience as dance, music, theater and art.

Factor IV was called Public Opinion since it was defined by five of the ten Public Opinion items from FJC IV (see Tables 3, 4). Of the remaining five items, one item formed

a single item factor and was therefore deleted, two items each clustered to form Factors VI, Personal Judgment and Factor VIII, Public Opinion: Personal Involvement. It will facilitate the discussion to describe Factors VI and VIII before commenting on Factor IV.

Factor VI was labeled Personal Judgment since the two items defining this scale reflect an individual's general (positive) attitude towards black people.

Factor VIII was named Public Opinion: Personal Involvement because the two items defining the factor reflect a specific-personal aspect of public opinion.

The items defining Factors IV, VI and VIII were judged in the LPA study (see Table 3) to measure Public Opinion (LC 4) and Personal Judgment (LC 10). The items defining these two LC's were grouped by us to form the FJC IV, Public Opinion. The appropriateness of this item grouping can be examined in light of the interrelationships uncovered among items using factor analysis. Consideration of the discrepancies between the judgmental and response data adds information to the interpretability of the intended LC, Public Opinion. For example, the judges tended to categorize the items reflecting rather general opinions of whites toward black people, a specific individual's opinion toward a black person as well as opinions possibly reflecting more personal involvement, into a general Public Opinion category (FJC IV). But the response data dimensions suggest that interpretation of these public opinion items would be more meaningful in the context of the three aspects of public opinion reflected in Factors IV, VI and VIII.

Factor V, Philosophy, was clearly defined by five of the six items from FJC VI, Philosophy. It is interesting to note that agreement with the items of this dimension indicates a respondant's support of SCLC, CORE and NAACP but not the Black Panthers. The final Philosophy item dealing with the credibility of the Black Muslims did not contribute to naming this factor.

Factor VII, Community Control, was defined by the four items from FJC V. Agreement with the item content reflects support of black community control of business, police, education and private industry.

Finally, Factor IX has been labeled Student Image because it was defined by three of the five items from FJC VII, Student Image which mainly reflected views of teachers regarding the classroom behavior of black students. The remaining two Student Image items, reflecting personal qualities of black students, loaded across separate factors which were not selected for discussion.

It should be noted that these data do not warrant response data clustering of the five items from FJC II, Black Studies Courses.

Comparison of LPA and Factor Analytic Data

In this section additional comments are made concerning relationships between the results of judgmental data and response data analysis.

With the exception of the Black Studies Courses items, substantial agreement was found between the response data factors and the judgmentally derived categories. Examination of discrepancies between the categories and factors (e.g. FJC IV Public Opinion and Factors IV, Public Opinion VI, Personal Judgment and VIII, Public Opinion: Personal Involvement) contributed to a better understanding of the attitudinal constructs under investigation. Consideration of the amount of agreement between the categories and factors for these data in light of the internal consistency reliabilities of the factors suggests that reliabilities tend, as expected, to be higher for those scales with more items and for the subscales which corresponded jointly in judgmental and response data clusters.

Finally, it should be noted that additional analyses of response data illustrated in this study were carried out, but not reported in this paper. A confirmatory method of factor analysis, Guttman's (1952) rank reduction procedure, was employed by Simons (1971) to empirically examine the response data after an a priori hypothesizing of the specific item clusters on the basis of the LPA judgmental data. Results of the analysis were generally supportive of the interpretations found in exploratory principal components analysis. Also results of an oblique rotation of the principal component loading matrix and an image analysis followed by both orthogonal and oblique rotations (Hofmann, 1970) were found to agree with the factorial interpretations which have already been reported.

Discussion and Conclusions

In this paper it has been proposed that objective methods (using LPA) may be employed to identify certain subuniverses within a universe of items. Given an appropriate selection of judges (sorters), items, and sorting directions, one might expect individual derived item categories to be predictive of ultimate response data categories. In turn, this strategy might be sequentially applied as item pools are modified, and as new response data is collected. This sequential process, may be regarded as a means to assist in construct validation since naming of one's ultimate item categories should be facilitated by an examination of relationships between judgmentally derived and response data derived clusters. As Cronbach (1971) notes, both item and test validation are fundamentally a process of interpretating test scores, and findings those interpretations for which an item (or test) is valid. There are strong parallels between attitude and achievement test development so distinctions between these types of tests seems not to be crucial. It seems reasonable further to examine the present methodology in relation to facet systems for item content specification although such a task will not be attempted here.

It is not suggested that the LPA method is necessarily sufficient for judgmental data analyses, but it does appear to fill a certain chasm in the spectrum of typical test developmental methods. If additional empirical studies of items in a content domain do show that there is a general correspondence between derived LPA categories and results of response data analyses, numerous further uses of the proposed strategy become obvious. Construction of "strictly parallel" forms and score equating across forms are cases in point. It must be reiterated, however, that such uses will require empirical study of relationships between results of categorical data studies and response data studies for the same items. Also, it may be quite important to select one's method for analyzing response data characteristics.

Some comments on sampling problems related to LPA studies are also necessary. Three fundamentally different types of sampling are involved in any judgmental study; one must select items, sorters and sorting instructions. An investigator will, of course, be prudent to restrict his attention to a pool of high quality items (those for which agreement exists that some relevant characteristic may be measured). Results of a judgmental study may be meaningful explicitly only within the context of the particular items chosen. Items should not be overly complex in form, nor should they usually contain more than one principal idea. To the extent that items are highly complex, the likelihood for interpretable results is apt to be substantially diminished.

The initial selection of persons to sort the items can also be critical. Certainly some persons might produce categorizations greatly different from those of other persons. Sorters should probably be selected to have experience or competence in the subject-matter area if categories are to be most interesting. For some ideas on how one might study relationships among different persons' categorizations of the same items see Tucker and Messick (1963) and Pruzek, Stegman and Pfeiffer (1972).

The last problem to be considered is that of selecting instructions for sorting. Instructions in the present empirical study were designed with the intent of deriving categories which reflected content characteristics of the items. It is possible that different instructions, even if generally based on the principle of content differentiation, might have resulted in different derived categories. While such a result appears unlikely to these investigators; it remains a possibility. More generally, we can conceive of entirely different bases for category generation, e.g., item content complexity, which could be fruitful. All of these latter questions, as well as those relating to the sampling of persons and sorters, must be viewed as questions for further study.

References

Bloom, B. S. (Ed.). Taxonomy of Educational Objectives: Handbook I, Cognitive Domain. New York: David McKay Co., Inc., 1956.

Cronbach, L. J. Test Validation. In R. Thorndike (Ed.) Educational Measurement. Washington, D. C.: American Council on Education, 1971, 443-507.

Ebel, R. L. How an examination service helps college teachers to give better tests. Proceedings of the 1952 Invitational Conference on Testing Problems. Princeton: Educational Testing Service, 1953.

Johnson, S. C. Hierarchical clustering schemes. Psychometrika, 1967, 32, 241-254.

Lindquist, E. F. (Ed.) Educational Measurement. Washington, D. C.: American Council on Education, 1951.

McGuire, C. Research in the process approach to the construction and analysis of medical examinations. 20th Yearbook, National Council on Measurement in Education, 1963, 7-16.

Rozeboom, W. W. Foundations of the Theory of Prediction. Homewood, Ill.: Dorsey Press, 1966.

Simons, H. F. The Development of a scale to measure the attitudes of individuals towards black people. Unpublished Doctoral Dissertation, University of Connecticut, December, 1971.

Simons, H. F., and Gable, R. K. Measuring Attitudes Toward Black People. Paper and Symposia Abstracts 1972 Annual Meeting, Washington, D. C.: American Educational Research Association, 1972.

Stoker, H. W., and Kropp, R. P. Measurement of cognitive processes. Journal of Educational Measurement, 1964, 1, 39-42.

Tucker, L. Factor analysis of relevance judgments: an approach to content validity. Proceedings of the 1961 Invitational Conference on Testing Problems, Princeton: Educational Testing Service, 1962.

Tucker, L., and Messick, S. Individual differences in multidimensional scaling. Psychometrika, 1963, 28, 333-368.

Wiley, D. E. Latent partition analysis. Psychometrika, 1967, 32, No. 2, 183-193.

Appendix A

Component Loading Matrix Using A Varimax Rotation For Simons Data*

	COMPONENTS															
Item	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	-18	-10	1	1	0	16	8	4	-18	1	-4	17	24	0	60	-6
2	-6	10	12	-3	-4	12	-3	-2	-1	-6	4	4	-4	-3	-7	10
3	28	4	0	1	-1	1	-12	3	3	0	17	-1	-7	-10	16	-2
4	-9	14	-17	29	42	-4	-23	14	13	6	-24	-41	-14	35	2	1
5	-3	3	17	11	0	13	3	4	6	7	-10	-2	72	21	5	-6
6	-17	10	15	1	3	1	57	0	22	1	-5	-9	9	-1	-2	20
7	1	-8	-5	81	-4	0	3	-1	1	-2	2	-1	1	1	-2	-5
8	-16	3	78	-1	2	7	16	-3	-5	-1	-16	0	-5	6	-13	1
9	-1	15	1	-14	13	-50	-5	1	-1	-5	7	5	13	-41	-55	-16
10	35	16	-22	2	-15	16	-28	-7	-12	17	-5	12	-6	72	30	-15
11	1	82	7	-14	6	6	4	-7	1	2	10	3	2	3	-5	16
12	-17	-9	15	-10	5	1	12	7	5	-21	3	-10	17	38	5	-9
13	-5	-2	-2	0	74	-7	2	5	16	-7	-16	-4	10	16	-5	8
14	25	8	-6	0	14	-16	-2	-1	5	17	70	10	-13	3	4	-3
15	11	25	5	0	14	-2	73	-7	-5	-16	-4	3	12	6	12	5
16	22	-21	3	5	7	-3	-76	0	2	-2	-3	1	5	-12	13	17
17	3	-16	-6	81	3	10	6	-5	-2	-10	-5	-7	4	3	3	11
18	-3	12	20	-9	-6	12	14	11	-7	-3	-13	3	5	-2	4	1
19	-4	2	4	9	6	73	0	3	8	-16	-1	-5	18	-5	-3	-3
20	77	-16	-2	1	-13	-3	5	4	-2	5	9	7	-9	2	-2	-16
21	-3	65	19	-13	11	-8	25	0	-11	-1	-7	-4	10	7	-8	10
22	14	-34	7	13	6	22	-26	22	4	9	-7	30	15	40	12	10
23	1	-6	-13	5	-37	9	-3	9	-11	-1	74	5	1	-4	-5	6
24	11	-7	20	7	-4	-25	15	-25	6	-30	-3	-22	35	-1	8	-14
25	16	-3	-10	58	-6	1	-17	-15	11	13	26	-14	5	13	-2	8
26	-11	-1	81	-1	4	1	0	-7	2	-1	7	-16	9	2	0	5
27	-12	-4	10	-2	13	4	3	5	77	13	-6	-3	15	2	-9	6
28	49	9	-22	9	0	6	-22	10	-6	31	-1	-3	2	-32	-13	0
29	-2	63	6	-5	19	-14	12	11	-1	-3	-8	13	-10	9	4	-27
30	16	-15	-11	11	-6	-19	-4	-11	11	70	4	13	0	-5	0	1
31	16	-12	-6	13	-73	-12	-17	4	6	1	-10	11	7	17	-2	13
32	12	6	-10	2	-11	-10	-14	9	4	5	9	78	-10	03	7	13
33	-27	-11	5	53	-1	20	-6	24	14	25	-7	13	7	-14	11	-27
34	49	-2	-8	4	-3	12	-19	-6	9	30	10	16	-6	24	-11	-1
35	-2	43	6	-11	-32	-32	29	29	-4	21	-6	-3	5	3	5	17
36	67	1	-23	12	-7	-24	-11	9	-15	12	10	3	5	-6	-1	12
37	37	-5	-22	17	-36	-4	-29	7	1	7	16	-27	1	-1	-5	-14
38	11	-2	7	7	4	-1	-4	81	5	-7	-4	0	0	1	2	5
39	7	25	-1	-11	-3	2	4	7	-6	4	2	11	-11	-3	-5	75
40	-40	-5	3	-2	7	-4	7	7	17	-17	12	-11	45	-17	31	-8
41	64	-11	-17	5	2	7	-6	7	3	11	4	-2	5	-32	0	17
42	2	-20	9	70	-6	2	-7	27	-4	7	-5	15	1	-19	7	-25
43	7	52	-3	-11	1	-19	31	2	17	-18	3	-10	4	-25	10	-25
44	20	6	20	15	7	7	-5	12	24	-56	10	-11	-4	9	31	-8
45	-11	60	-1	-5	10	2	26	13	6	-9	-32	3	4	-5	1	3
46	3	4	-1	6	1	7	7	-1	84	-10	2	5	-4	4	3	-12
47	-19	6	10	-25	23	20	25	20	-5	-11	-2	2	72	14	44	4
48	-15	16	56	3	-3	-7	-7	73	15	-19	4	3	15	17	29	-11
49	13	-20	-21	6	-19	8	-46	-5	-11	10	2	20	25	-14	-25	9
50	-3	13	63	-1	3	-11	-3	3	11	-7	0	2	4	3	2	-4

*All entries have been multiplied by 100. Note that factors 1-7 and 9-10 were the nine factors chosen for interpretation.

TABLE 1

Derived Approximation to Latent
Category Matrix for 51 Attitude Items
with 12 Categories*

Item Number	CATEGORY NUMBER 1	2	3	4	5	6	7	8	9	10	11	12
	1	2	3	4	5	6	7	8	9	10	11	12
10	89	0	2	9	4	1	-29	-4	-8	-4	7	30
12	110	2	-4	-1	0	1	-23	0	-19	0	26	5
13	117	1	1	0	-4	0	19	1	4	-2	-6	-21
14	102	-2	2	-5	0	-1	33	-1	0	-4	-17	-9
15	66	0	6	-1	1	0	-32	0	21	-3	-4	45
3	0	104	0	-1	-1	-1	0	1	1	0	-1	0
4	5	100	2	0	-2	0	-3	-7	13	-4	-4	0
5	0	103	0	-2	4	0	-3	2	-7	3	-2	3
6	-2	103	-2	0	0	0	-3	-2	0	-1	7	1
21	4	86	-6	2	-3	-1	14	7	-5	9	5	-7
39	0	103	0	-2	4	0	-3	2	-7	3	-2	3
33	6	-15	104	-9	11	-2	-2	8	-5	10	-6	-2
34	0	0	102	0	-1	0	-7	1	0	2	2	1
35	-3	2	99	-17	0	0	-1	0	1	15	4	2
37	12	-1	64	33	17	-3	-2	10	-3	-16	-15	-4
38	-1	1	96	1	-7	2	15	-4	22	-13	1	-11
46	-11	18	58	22	0	4	26	-4	-19	0	-1	8
48	0	7	98	0	-10	1	15	-6	16	-10	-1	-9
49	1	-3	-3	108	-1	0	-1	1	-1	1	-1	0
50	1	-3	-3	108	-1	0	-1	1	-1	1	-1	0
51	-8	5	0	90	-3	0	5	-4	10	1	7	-2
29	0	0	0	-1	107	0	-1	0	-5	3	-1	2
30	0	0	0	-1	107	0	-1	0	-5	3	-1	2
31	0	0	0	-1	107	0	-1	0	-5	3	-1	2
32	-7	-1	-6	6	60	3	11	0	59	-20	8	-16
7	5	3	0	3	-4	93	8	-3	6	-6	1	-8
43	-1	-1	3	-5	-2	97	-1	5	-1	0	-2	1
44	-1	-1	-1	0	2	105	-2	-1	-2	0	0	3
45	-1	-1	-1	0	2	105	-2	-1	-2	0	0	3
17	28	-6	-14	-6	4	-7	50	0	41	21	32	-49
22	-3	-1	13	-1	-1	0	119	0	-12	-2	-6	-4
23	27	-3	-15	0	0	-1	97	1	12	-1	-21	7
24	-1	-9	-11	1	2	0	71	5	-28	11	39	24
25	-21	11	14	1	-1	1	86	-9	-10	-3	12	16
18	-7	-3	3	1	3	0	-3	101	8	-2	-6	6
19	-6	0	2	0	-2	0	-4	105	5	0	-5	7
20	14	2	-5	0	-2	0	8	99	-14	3	11	-13
1	13	1	-13	5	15	4	14	-3	76	-17	19	-15
8	-11	-3	14	8	-11	-2	-18	0	115	-4	-2	17
9	6	19	-6	5	3	13	6	23	47	-17	6	-16
16	-6	-3	1	-21	-13	-4	-3	1	87	52	8	6
11	22	0	1	9	1	1	-14	-1	1	71	4	-6
40	-10	2	0	2	2	1	0	0	8	99	-3	-1
41	-6	1	0	-1	5	0	2	0	-6	107	-2	-2
42	14	0	0	29	2	4	3	-1	-6	58	-8	3
2	0	5	-12	4	-4	0	-6	-6	33	-10	104	-5
28	32	1	11	-6	-2	0	21	-1	40	11	-48	43
36	-1	0	57	2	5	0	-31	1	-15	0	64	23
47	12	-8	9	-7	0	0	14	6	-32	13	85	6
26	-15	-6	-15	-1	3	0	36	7	9	-3	17	60
27	-10	5	-13	0	5	0	14	0	22	-1	6	69

*Rows were automatically reordered to facilitate interpretations;
All entries have been multiplied by 100.

TABLE 2

Indices of Association Between
Latent Categories for Simons LPA Results*

	CATEGORY NUMBER											
	1	2	3	4	5	6	7	8	9	10	11	12
1	77											
2	3	87										
3	-1	10	80									
4	8	3	10	83								
5	3	2	5	1	88							
6	0	8	1	0	17	87						
7	26	5	19	10	6	1	59					
8	7	48	0	0	17	11	5	89				
9	28	4	2	19	17	4	13	8	46			
10	16	0	3	58	-3	4	12	-1	26	90		
11	13	10	36	6	5	1	30	2	12	5	60	
12	63	0	8	12	-11	-3	54	-3	24	20	34	135

*Entries in this matrix will lie between zero and unity when the model fits the data. If the matrix is essentially diagonal, most manifest categories can be said to result from differential splitting of the same latent categories. Diagonal entries estimate the probability that any two items in that category will, in a new partition, be sorted into the same manifest category off-diagonal entries estimate the probability that two items from two different latent categories will be placed in the same manifest category.

TABLE 3

Summary of Original Derived Latent Category Labels,
Instrument Revisions and Final Judgmental Categories
Generated by LPA Procedure

Original Derived LC's	# of Items	Revisions Made	Final JC's		# of Items
1 Racism	4	combined with Bigotry; added 2 items			
2 Black Studies Courses	5	none	I	Racism	8
3 Inner City Teachers	5	combined with 1 item from Cultural Background; added 1 item	II	Black Studies Courses	5
4 Public Opinion	3	combined with Personal Judgment; added 5 items	III	Inner City Courses	7
5 Community Control	3	added 1 item	IV	Public Opinion	10
6 Philosophy	4	added 2 items	V	Community Control	4
7 Student Image	3	added 2 items	VI	Philosophy	6
8 Cultural Difference	3	added 2 items	VII	Student Image	5
9 Race Interaction	3	category deleted	VIII	Cultural Differences	5
10 Personal Judgment	3	deleted 1 item; combine with Public Opinion		TOTAL	50
11 Cultural Background	3	deleted 2 items; combined 1 with Inner City Teachers			
12 Bigotry	2	combined with Racism			
TOTAL	41				

*Note that the label LC 1-12 denotes the original derived latent categories whereas FJC I-VIII, denotes the final categories generated as a result of the LPA procedure.

TABLE 4

Factor Names, Original LPA Item Codes, Item Stems and Component Loadings for Derived Component Solution*

Factor	Original LPA Code	Item	Stem Summary	Loading
I Racism	R 13	20	members of black race have low moral standards	.77
	B 26	36	black students inarticulate & non verbal	.67
	B 27	41	black students apathetic, dull - seldom animated	.64
	R 14	28	majority blacks tend to be lazy-rather not work	.49
	R new	34	ethnic groups differ in innate ability	.49
	R new	40	blacks not lack individual initiative	.49
	Ph new	37	philosophy of Black Panthers not credible	.37
II Inner city Teachers	IT 34	11	no concern for welfare of blacks they teach	.82
	IT 33	2	concerned with salaries not needs of individual	.80
	IT 35	21	inner city schools-dumping grounds for inadequate teachers	.66
	IT new	29	do not involve themselves with community agencies	.65
	CB 36	45	rate behavior of black low even if achievement is satisfactory	.60
	IT 48	43	experience difficulty in understanding blacks	.52
	IT 38	35	teaching positions in inner city not attract able & ambitious teachers	.43
III Cultural Differences	CD new	50	there is black dance	.85
	CD 20	26	there is black music	.81
	CD 19	18	there is black theatre	.80
	CD 18	8	there is black art	.78
	CD new	48	there is black folklore	.56
IV Public Opinion	PO 49	7	whites at university level sympathetic to the problems of blacks	.81
	PO 50	17	whites at university level want equality for blacks	.81
	PO new	42	if blacks given equal opportunity, probably make success of lives	.70
	PO 51	25	university attempts meet black needs	.58
	PO new	33	whites in university believe if blacks given good job, make a success of lives	.53
V Philosophy	Ph 44	15	philosophy of SCLC is credible	.74
	Ph 7	31	philosophy of Operation Bread Basket credible	.73
	Ph 43	4	philosophy of NAACP is credible	.42
	Ph 45	23	philosophy of CORE not credible	.37
	Ph new	37	philosophy of Black Panther not credible	-.36
VI Personal Judgment	PJ 41	19	when interact with blacks, not aware of skin color	.73
	PJ 40	9	if black were president of U.S., not forget skin color	.58
VII Community Control	CC 31	15	community control of education desirable black community	.73
	CC 30	16	community control of business desirable black community	.70
	CC 29	6	community control of police desirable black community	.67
	CC new	49	community control of private industry not essential in black community	.46
VIII Public Opinion: Personal Involvement	PO new	46	if newborn black baby adopted into black middle class family, learn middle class values	.83
	PO new	27	if newborn black baby adopted into white middle class family, learn middle class values	.77
IX Student Image	SI 25	30	way to maintain discipline is to be strict & direct	.70
	SI 22	3	black students discipline problem; toughness - comic activities	.66
	SI new	44	blacks generally civilized as whites	.56

*Note that for each item the original LPA category code and item number are listed for comparison with Table 1. For example, item 20 in Factor I: Racism was originally item 13 in the category labeled Racism in the LPA study.

TABLE 5

Factor Name, Number of Items Corresponding to LPA Suggested Categories, Number of Items Defining Each Factor, and Estimated Alpha Internal Consistency Reliabilities

	Factor	Number of Items Corresponding to LPA Suggested Categories	Number of Items Defining Each Factor	Estimated Alpha Scale Internal Consistency Reliability
1.	Racism	6	7	.76
2.	Inner City Teachers	7	7	.82
3.	Culture Differences	5	5	.84
4.	Public Opinion	5	5	.77
5.	Philosophy	5	5	.64
6.	Personal Judgment	0	2	.23
7.	Community Control	4	4	.71
8.	Personal Opinion: Personal Involvement	0	2	.47
9.	Student Image	3	3	.59